Full Length Article

Computer aided framework for designing bio-based commodity molecules with enhanced properties

V. Gerbaud [a,*], M. Teles Dos Santos [b], N. Pandya [c], J.M. Aubry [d]

[a] Laboratoire de Génie Chimique, Université de Toulouse, CNRS, INPT, UPS, Toulouse, France
[b] Departamento de Engenharia Química, Escola Politécnica da USP, Av. Prof. Luciano Gualberto, Trav. 3, 380, 05508-900 Sao Paulo, Brazil
[c] Birla Institute of Technology & Science, Pilani, Dubai Campus, Dubai International Academic City, P. O. Box no. 345055, Dubai, UAE
[d] Université Lille 1 and ENSCL, Unité de Catalyse et de Chimie du Solide, CNRS UMR 8181, Equipe CÏSCO, Cité Scientifique, F-59655 Villeneuve d'Ascq, France

HIGHLIGHTS

- We propose a decision-making framework for designing sustainable molecules.
- The process involves managers, business experts, chemists and chemical engineers.
- We use computer aided molecular design to design bio-sourced molecules.
- We combine computer aided organic synthesis and computer aided molecular design.
- We enhance properties of existing molecules by finding suitable radicals.

Keywords:
Sustainable design framework
CAMD
Bio-based molecules
Fatty acids
Aprotic dipolar solvent
Sustainable design process

ABSTRACT

We investigate the use of computer aided molecular design (CAMD) approach for enhancing the properties of existing molecules by modifying their chemical structure to match target property values. The activity of tailoring molecules requires to aggregate knowledge disseminated across the whole chemical enterprise hierarchy, from the manager level to the chemists and chemical engineers, with different backgrounds and perception of what the ideal molecule should be. So, we propose a framework that allows the search to be successful in matching all requirements while capitalizing this knowledge spread among actors with different backgrounds with the help of SBVR (Semantics of Business Vocabulary and Rules) and OCL (Object Constraint Language). In the context of using biomass as the feedstock, we discuss the coupling of CAMD tools with computer aided organic synthesis tools so as to propose enhanced bio-sourced molecules which could be synthesized with eco-friendly pathways. Finally, we evaluate the sustainability of the molecules and of the whole decision-process as well. Specific applications that concern the use of bio-sourced molecules are presented: a case of typical derivatives of chemical platform molecules issued from the itaconic acid to substitute N-methyl-2-pyrrolidone NMP or dimethylformamide DMF solvents and a case of derivatives of lipids to be used a biolubricants.

1. Introduction

Since its emergence in the 80's, Computer Aided Molecular Design has become a standard tool for finding molecules targeting desired properties (Gani and Brignole, 1983). As exemplified in the 2003 compilation book "Computer Aided Molecular Design: Theory and Practice" under the supervision of Achenie, Gani and Venkatasubramanian (Achenie et al., 2003), successes were achieved in many application fields, particularly in solvent design, zero-CFC refrigerant design and polymer design. The CAPD (computer aided product design) extension aiming at designing mixtures was proposed as some synergy effects may appear between mixture components to match more closely with property requirements. Specific solving strategies of CAPD problems have been devised to find then both the molecules and the mixture composition: sequential (Gani, 2004; Karunanithi et al., 2005; Conte and Gani, 2011; Conte et al., 2011; Papadopoulos et al., 2013; Samudra and Sahinidis, 2013; Mattei et al.,

* Corresponding author.
E-mail address: Vincent.Gerbaud@ensiacet.fr (V. Gerbaud).

2014a), global (Vaidyanathan and El-Halwagi, 1996) and evolutionary search strategies (Heintz et al., 2014a, Herring and Eden, 2015).

Then, stricter environmental, health and safety (EHS) regulations rose in the 90's and they set new pressure on chemical industry actors for finding cleaner, healthier, safer and yet performant molecules. Those molecules could well be the results of a CAMD search. Regarding the REACH (Registration, Evaluation, Authorization and Restriction of Chemicals) regulation launched in 2006 in EU (REACH, 2006), chemicals to be registered concerned in the pre-registration period that ended in Dec. 2010 those who were produced over 1000 tons/year in the EU for any substance and 1 ton/year for CMR (Carcinogenic, mutagenic and reprotoxic) substances. It has led to 143,000 pre-registrations. Since 2013, the limit has now dropped for all substances to a minimum volume of 1 ton/year in the EU. These chemicals are spread in all kind of processes concerning almost all the industrial domains. In 2012, the cost of registering chemicals to comply with REACH was estimated to exceed € 2.1 billion, based on about 30,000 substances (ECHA, 2012).

An interesting concomitant coincidence was the popularization in the late 90's of the 12 green chemistry (Anastas and Warner, 1998) and 12 green engineering principles (Anastas and Zimmerman, 2003) which has encouraged the use of biomass as raw material. This translates into R&D activities and some economists have argued that a doubly green chemistry perspective prevails among the chemical companies engaged in green activities: one perspective concerns the reduction of their impacts on environment and the other concerns the use of renewable raw materials (Nieddu et al. 2012). The first perspective is a direct transcript of the definition of sustainable growth presented by the Bruntland report (1987). The second is the seventh principle of green chemistry (Anastas and Warner, 1998) and would induce sustainable issues like toxicity (principle 4), degradability (principle 10) or safety (principle 12) to be met more easily. We have recently proposed to bias CAMD approach towards the search of molecules that may bear bio-sourced fragments selected by the user (Heintz et al. 2014a).

A drawback of CAMD approaches is that no safeguard ensures the viability of the generated structures. Therefore, the best structures proposed might be virtually excellent but practically infeasible on an industrial scale. Furthermore, one would like to use eco-friendly chemical synthesis pathways (Bandres et al., 2011) to produce new solvents. Indeed, Anastas and Warner's green chemistry principles number 2 (maximize raw product utilization), 3 (minimize toxic molecule use and release), 6 (keep energy consumption low by using ambient T and P), 8 (use direct synthesis pathways), and 9 (use catalysts) concern chemical synthesis. The usual answer to that challenge in CAMD, and more generally when scheduling the development of a new molecule, is to check after the CAMD step whether the molecules can be produced with an eco-friendly synthesis step (Bergez-Lacoste et al., 2014). An alternative approach to secure an eco-friendly synthesis was proposed by Moity et al. (2014). They have screened hundreds of chemical reactions and extracted a short list of 53 reactions that ranked high in terms of an aggregated DEF criterion (Durability, Easiness and industrial Frequency of use). A Computer-Aided Organic Synthesis (CAOS) tool named GRASS was developed. But thousands of molecules are generated by CAOS and it is difficult to screen out the least promising with respect to desired properties (Moity et al., 2014). This can be done with the help of a performance function (Moity et al., 2016) like those used in CAMD.

With CAMD generating a lot of non-viable structures and CAOS leading to a combinatorial explosion, we imagine in the present paper to solve that issue by generating with CAOS viable structures to be modified by CAMD (Moity et al., 2016). Therefore one could trace CAMD candidate molecule to renewable raw materials through the CAOS most performant candidate produce in an eco-friendly manner.

But running this sequence requires a framework broader than that of CAMD or CAOS to be successful. First, in practice, the chemical product development process involves chemical engineers (CAMD users) and chemists (CAOS users) but also many stakeholders across the strategic and tactical layers of the chemical enterprise. Properties related to product requirements are usually classified as essential, desirable and EHS properties (Harper and Gani, 2000) or similarly as product, process-related or usage-related (Costa et al., 2006). Typical EHS attributes are often decided at the strategic and tactical levels of the enterprise by managers, business process experts, product portfolio leaders, marketing office and legal department. They take the opportunity to develop their business in green chemistry either in response to a client demand or to changes in rules and regulations or any other stimuli from the outside of the company.

Second, considering the growing complexity of designing a new product that qualifies as sustainable, Hung et al. (2008) stressed the importance of information and knowledge management between people that are coming from different cultures and the importance of bringing them to consider all the sustainability issues together. This sets additional challenges. An evident solution is to wrap the chemical product design and development process into a decision-making process (Harper and Gani, 2000), following Simon's intelligence, design and choice phases (Simon, 1960) and to allow all stakeholders to be involved through a hierarchical procedure matching the enterprise organization levels such as proposed by Ng (2004) or Heintz et al. (2014b). Similar decision making approaches for integrating sustainability into process design have been proposed in the literature (Azapagic et al., 2006). But stakeholders from different culture would state requirements differently. Corporate managers will use words to express real needs (e.g. solubilizing) that should be translated for CAMD tools inputs in terms of computable property target values (e.g. solubility from SLE calculation, solubility from Hansen solubility parameters). Engineers will use mathematics, physics, chemistry and numbers to express them. In CAMD approach this translation can be achieved through problem templates (Mattei et al., 2014a, 2014b) or property relations based on user-defined attributes (Solvason et al., 2009). To harmonize the expression of requirements and their understanding by all stakeholders, Heintz and coworkers have proposed to use unambiguous semantic languages like Object Constraint Language (OCL) (OMG, 2006) for hard science statements and Semantics of Business Vocabulary and Rules (SBVR) (OMG, 2008) for literal statements (Heintz et al., 2014b).

In this paper we refine the framework of Heintz et al. (2014b) that describes the chemical enterprise wide process for finding new commodity chemicals with a CAMD approach combined with CAOS to ensure that new molecules are traceable to renewable raw materials through eco-friendly synthesis pathways. It could help chemicals manufacturers and resellers who oversee the production of a limited set of chemicals to imagine derivatives as substitute candidates. Section 2 describes the semantic languages SBVR and OCL, computer aided tools for molecular design and chemical synthesis and tools for assessing the sustainability of the whole design process. Section 3 describes the framework and exemplifies it through the solving of CAMD problems that concern the finding of bio-sourced molecules, typically derivatives of lipids and derivatives of chemical platform molecules issued from the itaconic acid feedstock as substitute solvents.

2. Materials and methods

2.1. SBVR and OCL language

With the involvement of many stakeholders coming from different cultures, we need to describe requirements unambiguously.

The Semantics of Business Vocabulary and Rules (SBVR) standard (OMG, 2008) is a language that allows expressing unambiguously so-called business rules that describe rules in use inside the enterprise. For example, consider the following rule expressing the need of a customer:

- Fact1: **Customer C1** ***wants*** **a replacement product that** ***is greener than*** **product solvent A.**

Red terms are keyword terms for modality, the underlined terms designate standard and specific objects, and italic terms designate facts and verbs. No maths or values are usually described. Here we use SBVR for expressing business rules at the strategic and tactical levels of the enterprise stated by the manager, business expert and product portfolio manager stakeholders that are involved to define together requirements of a new sustainable product.

The Object Constraint Language (OCL) (OMG, 2006) is another standard language allowing the description of numerical constraints and other attributes of an object as in the following example that describes an imaginary aqueous product to be designed with water imposed as the first molecule:

Context Molecule **inv**

self.molecule[1]- > type()="FixedMolecule"
self.molecule[1].list- > size()=1
self.molecule[1].list[1].MoleculeName={water}
self.molecule[1].list[1].MoleculeFormula={H_2O}

Here we use OCL at the operational level of the enterprise to help chemists and chemical engineers stakeholders to formulate constraints that specify requirements of a new sustainable product: synthesis pathways, biomass feedstock, co-reactants, target properties, CAMD building blocks. Both languages are exemplified later.

2.2. Computer Aided Organic chemical Synthesis tool: GRASS

Moity et al. (2014) selected 53 sustainability-compliant synthesis pathways with high DEF (Durability, Easiness and industrial Frequency of use) and implemented it in the GRASS Software (GeneratoR of Agro-based Sustainable Solvents). This Computer-Aided Organic Synthesis (CAOS) software derives from GRAAL, a software developed by Barone et al. to generate all possible degradation products of food flavors (Barone et al., 2005). The detailed description of the program architecture is given in the original papers (Barone et al., 2005, 2010). It requires as input a bio-sourced building block, a list of readily available co-reactants and a list of eco-friendly chemical transformations suitable for large-scale productions. To limit the unavoidable combinatorial explosion, only one representative of each family of the co-reactants is used (eg. methanol for alcohols, acetone for ketones, and methylamine for amine …). Implicit co-reactants also considered are H_2, H_2O, H_2O_2, CO, CO_2, NH_3, O_2, O_3, SO_3, ethylene, ethylene oxide, formaldehyde. Then, all possible products for each transformation and co-reactant are enumerated. The outputs are the generated structures which are encoded by their SMILES code. Typically, thousands of structures are produced at each generation.

In our first case study, we consider as a renewable raw material feedstock the itaconic acid, listed by the US Department of Energy (Werpy and Petersen, 2004) as one of the Top-12 biomass synthons. Starting from that synthon, Moity et al. (2014) synthesized in-silico molecules with the help of GRASS, by using seven co-reactants with 53 selected chemical transformations. 40, $\approx$1000 and $\approx$17,000 compounds were generated by GRASS starting from itaconic acid after one, two and three steps respectively (Moity et al., 2014). A tremendous structure exploration was carried out to retrieve five molecular structures that have already been reported in the literature as solvents: 2-methyl-butanediol (2-MBDO), 2-methyl-gamma-butyrolactone (2-MGBL), 3-methyl-gamma-butyrolactone (3-MGBL), 3-methyl-tetrahydrofuran (3-MTHF) and N-alkyl-4-carboxypyrrolidinone ester (NACP). Then Moity et al. (2014) noticed that NACP COSMO-RS sigma profile qualifies this molecule as a non H-bond donor and electron-pair donor compound which would belong to the aprotic highly dipolar solvents cluster according to Durand's classification (Durand et al., 2011). Among aprotic highly dipolar solvents, only a few alternatives green solvents are known (Moity et al., 2012). We intend to verify in the present paper with the help of CAMD approaches that NACP could extend the green solvents list suitable to replace toxic dimethyl formamide (DMF), N,N-dimethylacetamide, N,N-methyl-pyrrolidone (NMP) or less harmful dimethyl-sulfoxide (DMSO). Indeed NACP can be obtained in two eco-synthesis steps starting from itaconic acid, in agreement with a patent (BASF, 2009). This solvent family has also already been described as cosmetically acceptable (Richard and Muller, 2010) i.e. harmless to human health.

2.3. Computer Aided Molecular Design tool: IBSS

In a previous work we have conceived a novel Computer Aided Product Design (CAPD) tool, named IBSS (Heintz et al., 2014a). It follows CAMD precepts by first defining a set of target values for selected properties and by searching for new molecules – individual or mixtures – that best satisfy them. The problem solved by IBSS is a multi-objective simultaneous search over the optimization variables that are the

mixture molecules described by a molecular graph, the mixture composition and the mixture operating conditions. A genetic algorithm is used with specific genetic operators which are detailed in Heintz (2012). The candidate solution structure is imposed beforehand, namely the number of molecules in the mixture, the number of fragment in each molecule, and the composition limits. The performance of each candidate molecule or mixture is calculated from a performance function that will be described in the next section and which takes into account deviation of predicted property values from the target values.

IBSS can generate molecular fragments "from scratch" from selected building-blocks as any CAMD tool or bias the search towards specific molecules structures where some fragments are imposed. This allows, in our case, to select fragments that can be traced from the biomass feedstock, like the NACP fragment that was provided by the GRASS software aforementioned. IBSS input file follows an in-house extended markup language (XML) scheme describing literally all molecule structure constraints, building blocks, property models and target values while being readily translated into the object oriented package that feeds IBSS search. Output files follow an XML format based on the aggregation of standard chemical XML like CML (Murray-Rust and Rzepa, 2001) and property related XML like ThermoML (Frenkel et al., 2004). More details about file formats, software structure and functionalities are available in earlier works (Korichi et al., 2008; Heintz et al., 2014a).

In the first case study, we couple the CAMD search based upon biomass sourced fragments that are proposed by CAOS. Thereby we ensure the viability of the structures generated by the CAMD, in particular in terms of industrial production with eco-friendly chemical synthesis pathways.

2.4. Evaluation of molecule candidate performance

The fitness of each candidate molecule, represented by a molecular graph (MG_i) with respect to np target properties is computed with the help of Eq. (1)

$$Perf(MG_i) = \frac{\sum_{p=1}^{np} w_p \cdot F_p(MG_i)}{\sum_{p=1}^{np} w_p} \quad (1)$$

where w_P is the weight of each individual property.

Each property performance function F_p takes the value 1 if the target value is met. Deviation of property value P from target x is described by the following Gaussian type formula:

$$F_P(MG_i) = G(\text{Tol}, \text{val}) = \exp\left[\ln(\text{val}) \cdot \left(\frac{P - x}{\text{Tol}}\right)^2\right] \quad (2)$$

where the Tol (for Tolerance) value is the deviation from target at performance equal to the val parameter. For a given val, a small Tol means a rapid decrease of the performance. Tol values could be assigned to known standard deviation of property prediction models .

2.5. Evaluation of the sustainability of the design process

Gagnon et al. (2012) reviewed several conventional and sustainable design processes and proposed four stages to classify integrated sustainable engineering design process activities split in 22 tasks. Usual steps well known in CAMD approaches like design specification, alternative concept generation, alternative concept performance evaluation, are complemented with steps that Gagnon et al. claimed to be critical for achieving sustainability: multidisciplinary project team, definition of sustainability (sustainable issues relevant for the problem, criteria, assessment methods, monitoring indicators), selection of an approach for multi-criteria decision making, performance assessment according to the sustainability criteria or indicators. Finally, they proposed to evaluate the extent of sustainability achieved, by using four degrees of shade over six dimensions of sustainability ranging from A (minimal) to D (state of the art). The six dimensions cover the design process itself, the indicator relevance, the sustainability issue covered, the analysis tool accuracy, the alternative performance and the decision-making process itself. A case study will be analyzed in that way.

3. Results

We first propose the decision-making framework and second we exemplify it through the solving of CAMD problems that concern the finding of bio-sourced molecules with enhanced properties from itaconic derivatives and from a fatty acid.

3.1. Decision-making process for sustainable chemical synthesis and product design

In order to incorporate the possibility of biasing the CAMD search towards molecules that can be produced from renewable materials, we propose to monitor sequentially the decision-making process with four phases described in Fig. 1 as striped light brown boxes (shaded), starting from a stimulus (top left of the figure). Although it is not sufficient to handle all business decision variables recorded by Smith and Ierapepritou (2010) for chemical product design, it encompasses reasonably well the chemical product design activities before their implementation (Hill, 2009) and enables all stakeholders (light blue boxes) to get involved as recommended for a sustainable engineering design process (Azapagic and Perdan, 2005; Gagnon et al., 2012). Activities are highlighted in red over green rounded rectangles, while files and documents are assigned papersheet symbols.

The decision-making process follows Simon's classical three sequential phases (Simon, 1960), to which is added a knowledge

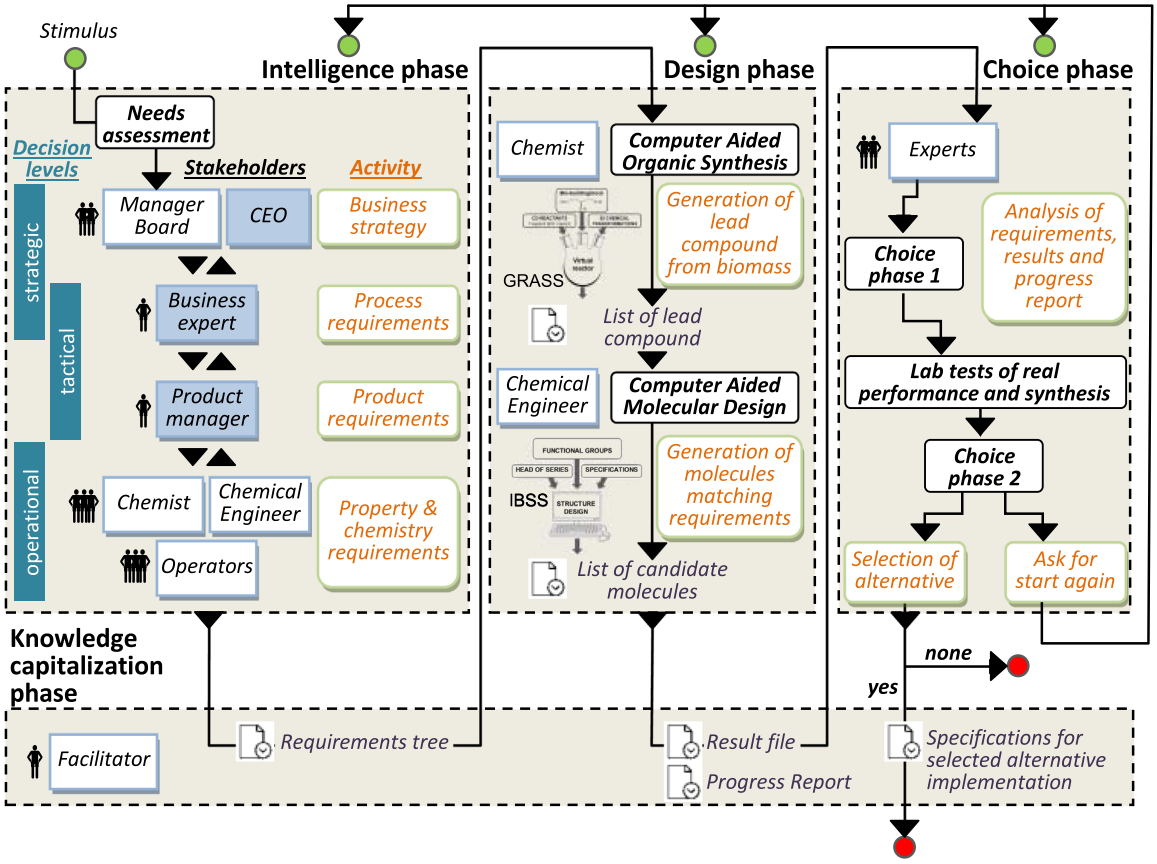

Fig. 1. Decision-making process for the design of molecule satisfying desired properties and being synthesizable from biomass.

capitalization phase supervised by a facilitator. The decision process is triggered by a stimulus that either conflicts with the business policy of the enterprise, like product substitution of a chemical to comply with the REACH regulation, or motivates a change in the business policy, like adding greener solvents to the product portfolio. Ultimately it ends with the selection of a candidate molecule to be later produced. Process design and implementation phases are not incorporated in the proposed frame.

3.1.1. The knowledge capitalization phase

During each phase, information is gathered and documents are produced. Hence, a usually implicit knowledge capitalization phase is made explicit in Fig. 1. This phase is assigned to a facilitator, which to our experience is a key person in running the full process successfully. Being in contact with all stakeholders, he should act as the project manager and be proficient with SBVR and OCL languages used in the intelligence phase during the building of a requirement tree, with all XML formats necessary for using the design phase tools and with the analysis of the experts' questionnaire answers in the choice phase.

3.1.2. The intelligence phase

The intelligence phases aims at gathering the chemical product requirements that meet customer needs and comply with the enterprise business strategy. We suppose that stakeholders across three levels of the enterprise pyramid are involved sequentially, concerned with decisions at the strategic (manager and CEO), tactical (business process expert and product portfolio manager) or operational level (chemical engineers, chemists and operators). The propagation of the decision across the enterprise levels ensures a strategic alignment as the decisions taken have an impact on the lower levels. This way the business strategy is followed at each level of the enterprise. We decide that requirements are expressed as business rules with SBVR at the strategic and tactical level, whereas the operational level stakeholders can directly express more quantitative constraints which are modeled with OCL. In addition, we use a coloring of the requirements to stress their interest priority, with status yellow for a “consider in priority” interest, white for a “consider as well” interest while shaded gray refers to “ignore” interest. The “ignore” status means that the requirement is kept in mind but is inactive for the current level. It may be reactivated along the construction of the requirements tree by other stakeholders, as all of them can access the entire tree details and documentation. A requirements tree building is detailed in the case study 1 below.

3.1.3. The design phase

The design phase aims at designing a new solution as a pure compound or mixture, possibly bio-sourced and hopefully with molecular structures viable in terms of chemical synthesis. It consists in using a sequential combination of Computer Aided Organic chemical

Synthesis (CAOS) and Computer Aided Molecular Design (CAMD) approaches as we exemplify later. GRASS and IBSS tools described before are used for the CAOS and CAMD steps respectively. Alternatively to GRASS, so-called lead compounds can be used as input to IBSS. In use in drug design, the lead compound term defines a chemical compound that has a property or activity likely to be useful, but that may still have suboptimal chemical structure that requires modification to fit better to the target.

The requirements tree data from the intelligence phase are used as input of the design phase. They are transformed into input datasheets for GRASS and XML product specifications for IBSS. GRASS output provides in particular the molecular structures of the lead compounds that will be set as fixed fragments in the molecule structure to be built by the CAMD tool. Some requirements like the cost evaluation, are unsuitable for IBSS, and are put aside to be considered later during the choice phase.

3.1.4. The choice phase

The choice phase consists in choosing which alternatives can be implemented. We select the DELPHI method (Dalkey and Helmer, 1963) for helping with the choice of the best alternative. In this method, experts are given access to all documents, including the requirement tree, are asked to answer a questionnaire. Results are analyzed by the facilitator. If no consensus is reached, a second round starts and the experts answer the questionnaires once again until the group converges toward a consensus answer. A questionnaire for chemical product design is available in the appendix of Heintz (2012). It contains four distinct sets of questions, about the performance of the alternatives (relevance of property choices, of property weighting, of target values, of property estimation methods), about the product structure (synthesis feasibility of each component of the mixture, of the mixture as a whole, mixture composition, potential oxidation, water stability and reactivity of the chosen molecule, the relevance of the fragments building blocks number and type), about the search algorithm (relevance of the algorithm parameter choices) and about general issues (large enough number of alternatives to compare, selection of the five best alternatives, specific rejection criteria and justification for some alternatives). If experts are not able to answer questions they can leave a “no opinion” answer.

The choice phase may end successfully or not as seen in Fig. 1 or ask for a new round of the design process at any phase start. For example, a new round may occur if some discrepancy is found between the real and the predicted performance, questioning the relevance of the property model accuracy used during the CAMD approach as we experienced in a previous work (Heintz et al., 2014a).

3.2. Case study 1: replacement of aprotic highly dipolar solvents with itaconic acid derivatives

3.2.1. Sustainability context and stimulus for greener solvent design

We consider a chemical enterprise involved in commodity chemicals production, especially solvents like NMP and DMF widely used in many reaction and separation processes. Because those solvents are potentially hazardous for the health, they have been the focus of research for substituting them by using CAMD (Weis and Visco, 2010). The stimulus starts with the acknowledgment that green solvents classified among aprotic highly dipolar solvents are rare (Moity et al., 2012). DMSO is one alternative but is not produced in the company and company customers inquire about new alternatives as they anticipate stricter regulations that would force them to substitute toxic NMP and DMF in their processes. Besides, aware of the threat that lower solvent sales would induce on the company's turnover, the CEO and board of executives are considering investing into biomass transforming industries so as to secure renewable material feed stocks for producing greener solvents in the future.

A new fact is created in the business rule repository of the enterprise at the corporate level following the SBVR standard where red terms are keyword terms for modality, the underlined terms designate standard and specific objects, and italic terms designate facts and verbs:

- Fact1: Customer C1 *wants* a replacement product that *is greener than* product aprotic highly dipolar solvent.

In addition, we list other facts and business rules existing in the enterprise repository:

- Fact2: Each thing that *is wanted* by a customer *is* a customer need *of* that customer.
- BusinessRule1: It is obligatory that each customer need *of* each customer *is satisfied*.
- BusinessRule2: It is obligatory that each replacement product *performs* the functionality that *is performed by* the product.
- BusinessRule3: It is obligatory that each raw material’s availability *is secured by* the company.
- BusinessRule4: It is obligatory that each product *respects* the security rules.
- BusinessRule5: It is prohibited that a product *prevents* a process *to function*.

Fact2 describes a customer need. BusinessRule1 shows that the commercial policy of the enterprise is to satisfy the customer needs.

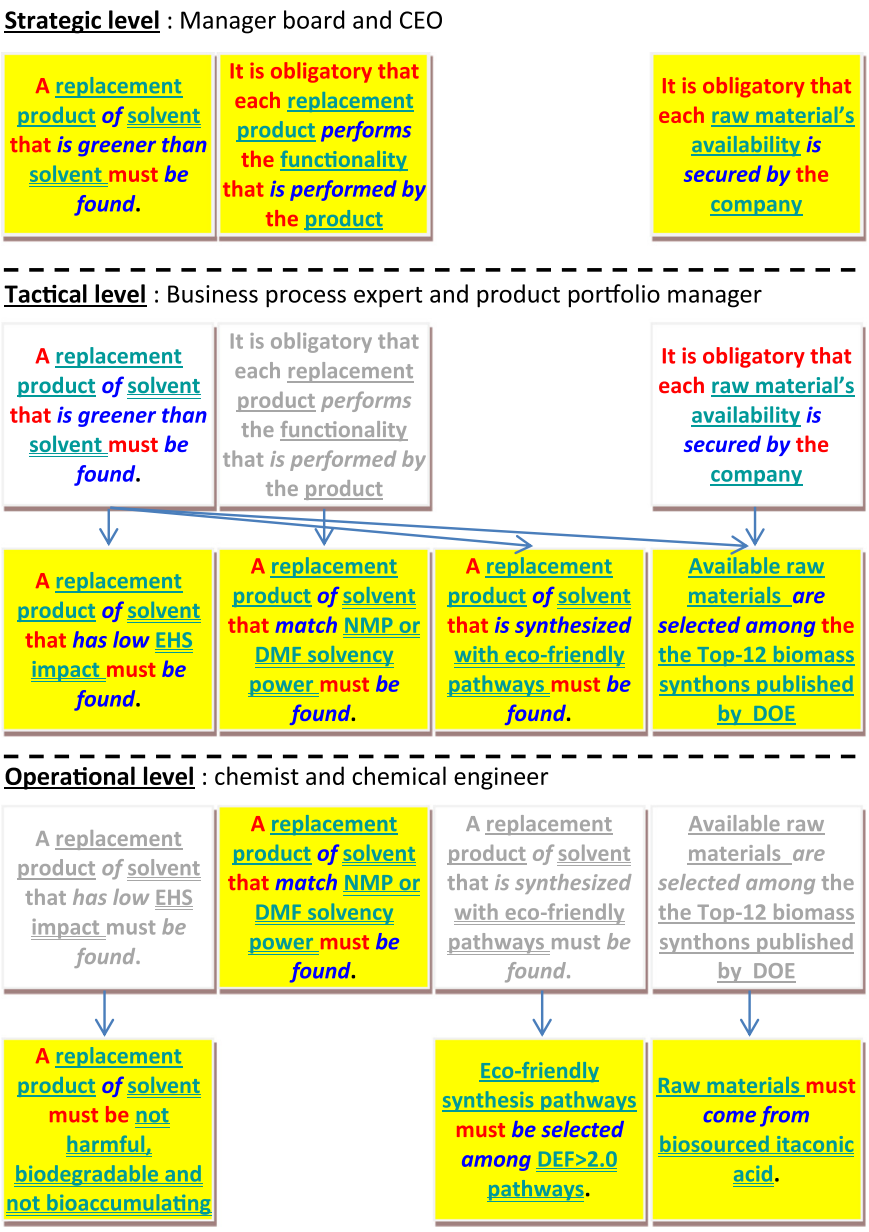

Fig. 2. First set of strategic, tactical and operational Business rules for the construction of the requirements tree for the design of aprotic highly dipolar greener solvent.

BusinessRule2 ensures that substituting products achieve the same key functionality (here matching the current solvent power of existing solvents). BusinessRule3 translates the company's policy in terms of securing raw materials suppliers. BusinessRule4 and BusinessRule5 refer to security and performance requirements.

The conjunction of Fact1 and Fact2 violates BusinessRule1: the customer wants a new solvent (fact1), which becomes a customer need (fact2) and as this is not currently satisfied, thus BusinessRule1 is violated The other BusinessRules are not violated but are cited for consideration since they may become violated by the substitution product.

3.2.2. Aprotic highly dipolar solvent intelligence phase

Induced by the business rule violation, the intelligence phase aiming at defining the requirements tree is initiated by the project manager, who takes the role of facilitator.

Figs. 2 and 3 describe business rules at the strategic, tactical and operational levels for the construction of the requirements tree for the design of aprotic highly dipolar greener solvent.

3.2.2.1. Business rules. At the strategic level, the project manager translates the enterprise business rules 1–5 into strategic local business rule 1, 2, 3, 5 and 6 expressed in SVBR (Fig. 2). Local business rules refer to temporary rules that apply only to the current decision making process. As this is the initial stage of the decision process, all rules are assigned a yellow color, meaning a "consider in priority" interest.

For conciseness, "solvent" is used from now in place of "aprotic highly dipolar solvent" in business rules. The enterprise BusinessRule1 becomes:

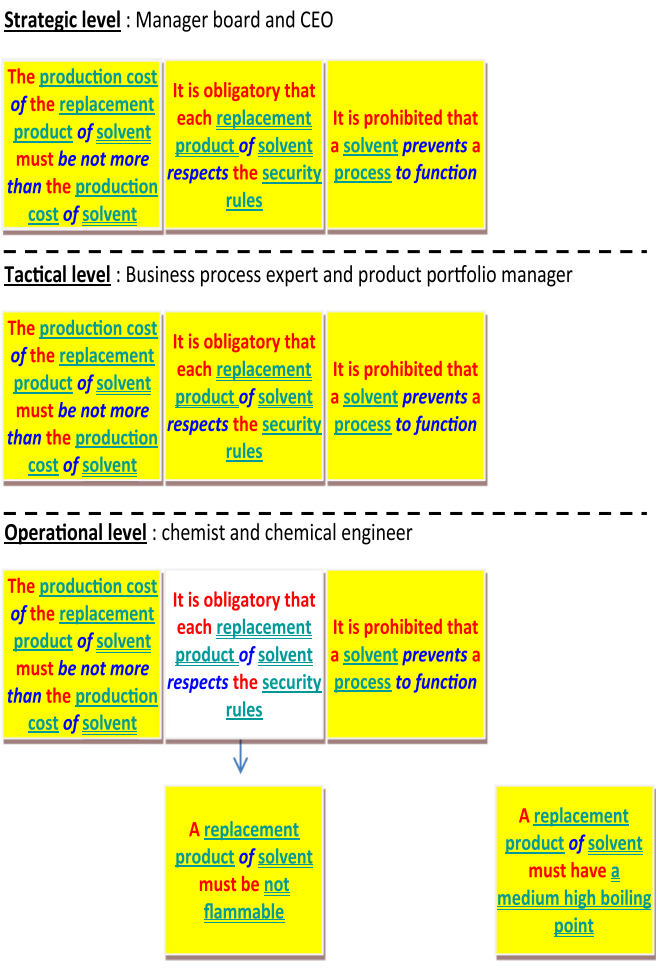

Fig. 3. Second set of strategic, tactical and operational Business rules for the construction of the requirements tree for the design of aprotic highly dipolar greener solvent.

- StrategicLocalRule1: A replacement product *of* solvent that *is greener than* solvent must *be found*.

A new StrategicLocalRule4 about production costs objectives is inserted by the manager.

- StrategicLocalRule4: The production cost *of* the replacement product *of* solvent must *be not more than* the production cost *of* solvent.

At the tactical level, the business process expert refines the StrategicLocalRule1 into TacticalLocalRule1 and TacticalLocalRule3 that orients the search towards a new solvent being not harmful, not flammable, biodegradable and not bioaccumulating, and imposes the chemical synthesis to be eco-friendly.

- TacticalLocalRule1: A replacement product *of* solvent that *has low* EHS impact must *be found*.
- TacticalLocalRule3: A replacement product *of* solvent that *is synthesized* with eco-friendly pathways must *be found*.

Then, StrategicLocalRule1 and StrategicLocalRule3 can be refined with the help of the product portfolio manager aware of market availability of raw materials to orient the search towards derivatives from biomass.

- TacticalLocalRule4: Available raw materials *are selected among* the Top-12 biomass synthons published by DOE

In the meanwhile the product manager refines StrategicLocalRule2 about functionality and assigns the goal that the new solvent should match existing in-house solvents NMP and DMF solvent power.

- TacticalLocalRule2: A replacement product *of* solvent that *match* at least NMP or DMF solvency power must *be found*.

The interest status coloring of each rule is also eventually changed: some strategic level rules are kept in yellow priority interest; new rules arise also in yellow state. StrategicLocalRule3 remains white because some economic issues about the securing of suppliers are not handled by the choice of biomass synthons expressed by TacticalLocalRule3. On the other hand, StrategicLocalRule2 status has shifted to gray “ignore” since the new TacticalLocalRule2 fully replaces it.

- OperationalLocalRule1: **A replacement product *of* solvent must be not harmful, not flammable, biodegradable and not bioaccumulating**

He also suggests to target a normal boiling point range matching NMP (BP 202°C) or DMF (BP 153°C). Those are high boiling solvents, used for this reason in separation processes like extractive distillation. This creates a new rule:

- OperationalLocalRule2: **A replacement product *of* solvent must have a medium high boiling point**

At the operational level, a chemical engineer stakeholder should refine TacticalLocalRule1 to detail the properties to be matched.

Besides, a chemist stakeholder would propose to select eco-friendly chemical synthesis pathways within the list published by Moity et al. (2014) with index DEF > 2.0. He also selects the itaconic acid platform molecule as raw materials for the current design process in dialog with the business process expert. This expert has indeed found some itaconic acid supplier which would accept participation in the company's assets and partial control, to secure the raw material feedstock as recommended by BusinessRule3. TacticalLocalRule3 and TacticalLocalRule4 shift to “ignore” status as they are then replaced by OperationalLocalRule2 and OperationalLocalRule3:

- OperationalLocalRule2: **Eco-friendly synthesis pathways must *be selected among* DEF>2.0 pathways.**
- OperationalLocalRule3: **Raw materials must *come from* biosourced itaconic acid.**

The business rules above lack precise statements about the type of chemical building blocks, the calculable property, the numerical value and the unit of each property. These data are required as inputs to the design phase. Hence, at the operational level, all yellow and white status business rules are translated into property target values constraints expressed in OCL (Figs. 4 and 5). Help from the facilitator is needed because the OCL constraints should contain all the information needed to define input files to the CAOS and CAMD tools used in the next design phase.

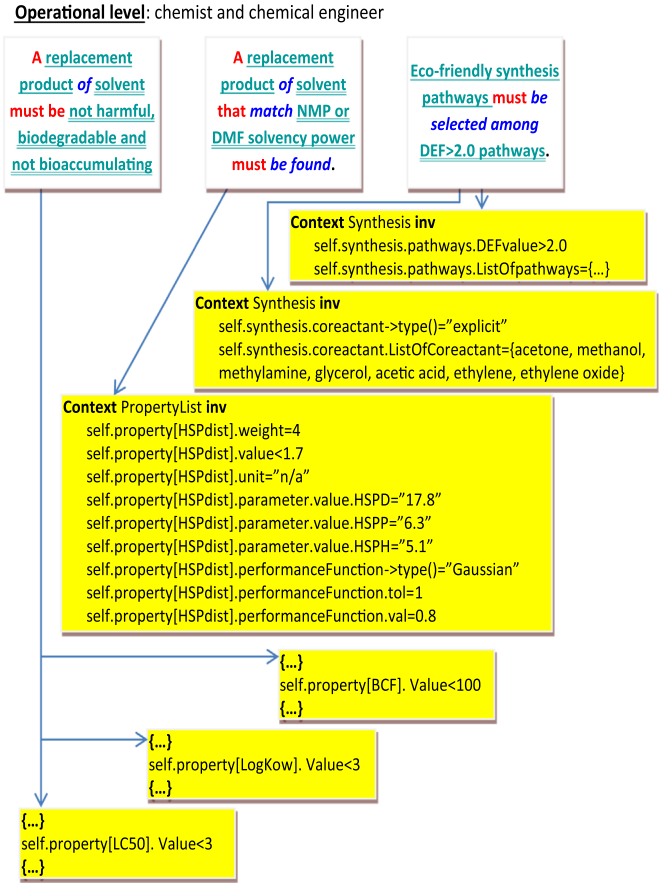

Fig. 4. First set of OCL constraints refining for the construction of the requirements tree for the design of aprotic highly dipolar greener solvent.

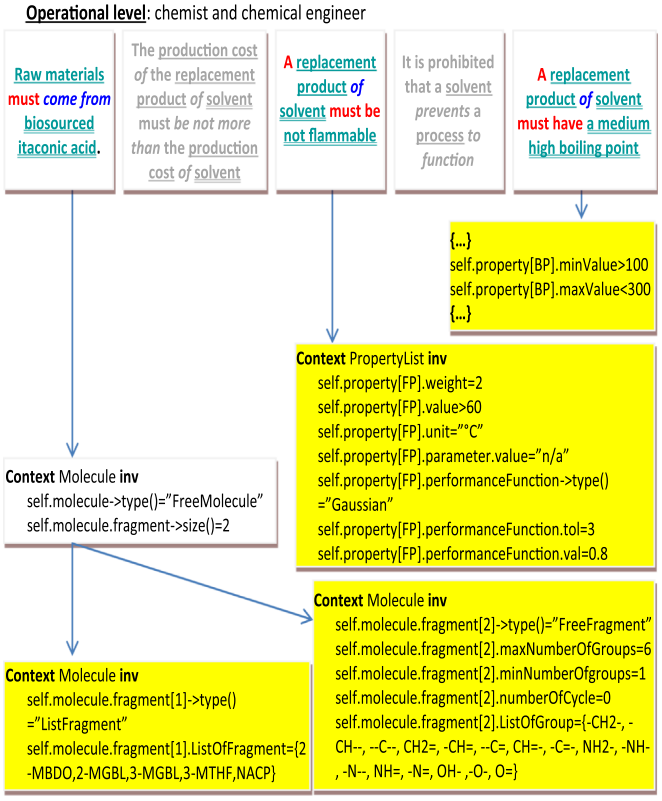

Fig. 5. Second set of OCL constraints refining for the construction of the requirements tree for the design of aprotic highly dipolar greener solvent.

Table 1
Calculable properties and models for the computation of molecule performance.

Real property	Calculable property	Target value	Prediction method	Performance function (Eq. (2))	Weight
Solvency power	HSP distance (Eq. (3))	1.7 for NMP and DMF, 3.6 for DMSO (see text)	MB2010	G (1;0.8)	4
Medium high Boiler	Boiling point	100–300 °C	MG2001	G (3;0.8)	2
Non flammable	Flash point	> 60 °C	CPN2006	G (3;0.8)	2
Low toxicity	LC50 (mol/L)	< 3	MY2001	G (1;0.8)	0.5
Biodegradable	Log K_{ow}	< 3	MG2002	G (0.3;0.8)	0.5
Low bioaccumulation	BCF	< 100	VK1975	G (1;0.8)	0.5

*MB2010:HSPIP Y-MB (2010); MG2001:Marrero and Gani (2001); CPN2006:Catoire et al. (2006); MY2001:Martin and Young (2001); MG2002:Marrero and Gani (2002); VK1975:Veith and Konasewich (1975).

3.2.2.2. Calculable property assignment. Properties so far suggested are so-called real properties and are not quantitatively specified nor related to any calculable property model. These must be specified at this point. Table 1 shows the translation of real properties into calculable properties along with the selected property models, target values, deviation functions and property weights. Property weights are assigned after questioning all stakeholders. But, overweighting inaccurate models in the performance function would lead to poor results. Candidate molecules ranked top would be screened out after experimental verification, which could be time consuming and costly. Alternatively, some good molecules would be missed. So, weights are lowered for LC50, log K_{ow} and BCF because the property models are not state of the art and their accuracy is questioned by the experts, as one would find that reputedly toxic NMP and DMF compounds are predicted mildly toxic by the chosen LC50 method. We keep these models for illustration purpose and because they will be questioned in the choice phase. It is also anticipated that experimental EHS data will be requested and much needed during the choice phase.

The Hansen solubility parameters (HSP) evaluate the solvency power of molecules (Hansen, 2007). Therefore, new solvents should match HSPs of existing solvents that they could substitute. Eq. (3) gives the solute–solvent distance D in the Hansen space

$$D = \sqrt{4(\delta_{d_{\text{Solvent}}} - \delta_{d_{\text{Solute}}})^2 + (\delta_{p_{\text{Solvent}}} - \delta_{p_{\text{Solute}}})^2 + (\delta_{h_{\text{Solvent}}} - \delta_{h_{\text{Solute}}})^2} \quad (3)$$

HSP values have been measured and reported in the literature. An ideal substitute would have a D distance equal to zero. But for the CAMD search we rely upon the HSPIP-YMB model that predicts values deviating from experiments for NMP, DMF and DMSO. Hence we propose to account for the model uncertainty by assigning a D value below which one would consider that a new solvent matches the NMP, DMF and DMSO HSP respectively. Table 2 lists experimental and predicted HSP values for the three targeted solvents along with D value computed from Eq. (3). The distance values are then taken as distance target values for new solvents, accounting for model uncertainty.

Table 2
Experimental and predicted values for the NMP, DMF and DMSO solvents.

	Exp. [a]			Predicted with HSPIP v3 [b]			D
	δ_d	δ_p	δ_h	δ_d	δ_p	δ_h	
NMP	17.8	6.3	5.1	17.7	11.0	6.3	1.69
DMF	17.4	13.7	11.3	16.8	14.8	11.4	1.63
DMSO	18.4	16.4	10.2	17.9	19.7	9.3	3.54

[a] Hansen (2007).
[b] MB2010:HSPIP Y-MB (2010).

Stakeholders have requested additional requirements related to health, safety and security. Toxicity is evaluated by setting a limit value for the 96 h fathead minnow acute toxicity ($-\log_{10}$ LC50 < 3, in mol/L). This selected target value corresponds to LC50 > 100 mg/L for a molecule weighting 100 g/mol, labeled as category 4 i.e. not harmful for aquatic environment in the United Nation Global Harmonization System of Hazard Classification and Labeling (UNGHS, 2013). Biodegradability is evaluated from the K_{ow} octanol–water partition coefficient. Nonpolar compounds usually have values of log $K_{ow} > 4$ and polar compounds have values of log $K_{ow} < 1.5$. A target value log $K_{ow} < 3$ is set. As regards to the bioconcentration factor (BCF), a value greater than 500 L/kg indicates potential bioaccumulation (UNGHS, 2013). It is often correlated to the K_{ow} value (Veith and Konasewich, 1975). Typically, a value of $3 <$ log $K_{ow} < 4$ corresponds to $100 <$ BCF < 500. A target value log BCF < 2 is set.

3.2.2.3. OCL constraints. With the setting of property targets for future molecules, the intelligence phase can go on to specify formally the constraints with the help of OCL and refine the business rules. With the help of the facilitator, OCL refining is focussed on providing the necessary information for running the CAOS and CAMD steps in the design phases (see Fig. 1). The whole requirements tree process construction is presented in Figs. 4 and 5, starting from the business rules presented in Figs. 2 and 3. It leads to the definition of 10 requirements that can be used in the design phase.

Regarding CAOS, the chemist is qualified to sets the list of co-reactants and of chemical transformations that will be used within the CAOS GRASS tool. For coreactants, the OCL is the following (Fig. 4):

Context Synthesis **inv**

self.synthesis.coreactant- > type()="explicit"
self.synthesis.coreactant.ListOfCoreactant={acetone, methanol, methylamine, glycerol, acetic acid, ethylene, ethylene oxide}

Notice that these coreactants are short molecules and single representative of chemical families, except for alcohols where a mono-ol and polyol is used. Therefore, the compounds generated by GRASS from this shortlist of co-reactants will be regarded as a lead compound that can be further extended in chain length for instance by using CAMD. For the chemical transformation, recalling Moity et al. (2014) classification of eco-friendly chemical pathways, the chemist stakeholder chooses those with DEF index > 2.0.

For the CAMD step, the chemical engineering expert can formalize constraints with the help of OCL. For example, the TacticalLocalRule1 about non-flammability is refined into an OCL constraint on the flashpoint limit value (FP > 60 °C) which corresponds to Class IIIA combustible liquids for the replacement product (see Fig. 5). Other OCL constraints cover the properties listed in Table 1 and are displayed in Figs. 4 and 5.

The chemist expert is qualified to set the molecule structure in the CAMD problem. He decides that the future green solvent is made of several fragments. One will be a fragment chosen from a list of derivatives from itaconic acid, to be set later after the CAOS step in the design phase. We recalled in introduction that the CAOS step was already run for the itaconic acid (Moity et al., 2014). It showed that the 2-MBDO, 2-MGBL, 3-MGBL, 3-MTHF or NACP fragment could be candidate substitute for aprotic highly dipolar solvents. The related OCL constraint with these five compounds is displayed in Fig. 5.

The other molecule fragments will be of free type, built as an assembly of chemical building blocks chosen among simple chemical groups with carbon, oxygen and nitrogen (see the related OCL in Fig. 5). Some groups which may enhance health hazard like halogens or aromatics are excluded at first. Non aromatic heterocycles are also avoided because they are usually naturally existent in the bio-sourced fragments, as illustrated by the fact that four out of the five itaconic derivatives above exhibit five atom heterocycles.

3.2.3. Aprotic highly dipolar solvent design phase

Following the decision making process, the design phase is split into sequential CAOS and CAMD steps. The CAOS step is not detailed in this paper because we can exploit Moity et al. (2014) results: from the itaconic acid platform molecule, 5 solvents are retrieved as candidate solvents, namely 2-MBDO, 2-MGBL, 3-MGBL, 3-MTHF, NACP. Using Eq. (1), we can evaluate their performance with respect to an objective of substituting NMP, DMF or DMSO (Table 3).

A closer look at individual property performances shows that EHS properties are not discriminant because these five molecules are biosourced and bear inherently some low impact features. But the flash point is more discriminant and is predicted below target for 2-MGBL and 3-MGBL while NACP has a flashpoint predicted above target. Overall NACP is a good candidate for all three targeted solvents.

The CAMD step is now detailed only for the NACP candidate that Moity et al. deemed suitable to replace NMP, DMF and DMSO. For the CAMD step, the OCL format compiling into the XML format input file of the CAMD tool IBSS can be fully automated. As shown in Fig. 6, OCL displayed in Figs. 4 and 5 enable to set the molecule structure data with a bio-based fragment with two connected free fragments and the structure data of the performance function based on property predictions.

Search algorithm specific parameters data are added by the facilitator based on preliminary tests (Heintz, 2012). The search is run over 300 generations and is completed in less than 20 min. The maximum number of building blocks in the free fragments allowed is 6. Other

Table 3
Lead compound performance.

Lead compound		Performance for substitution with respect to		
Acronym	SMILES	NMP (%)	DMF (%)	DMSO (%)
2-MBDO	OCC(C)CCO	57.9	57.9	57.9
2-MGBL	C1(=O)CC(C)CO1	79.5	40.0	44.0
3-MGBL	C1(=O)C(C)CCO1	78.0	39.2	43.4
3-MTHF	C1CC(C)CO1	15.8	15.8	15.8
NACP	OC(=O)C1CN(C(=O)C1)	57.9	61.9	88.5

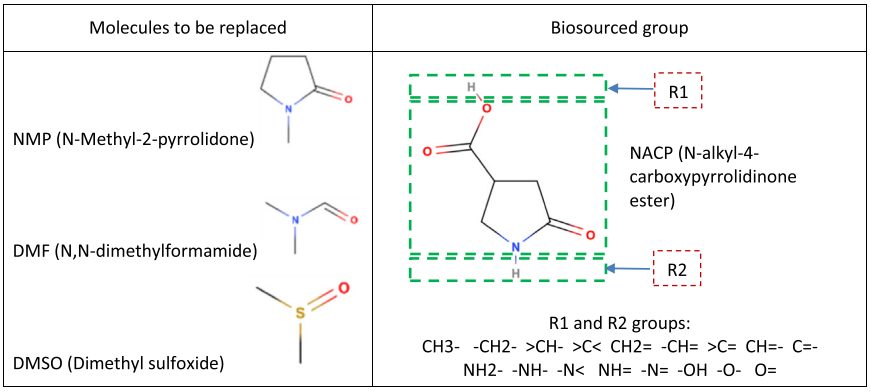

Fig. 6. Organic molecular structure and building blocks in the aprotic highly dipolar solvent search.

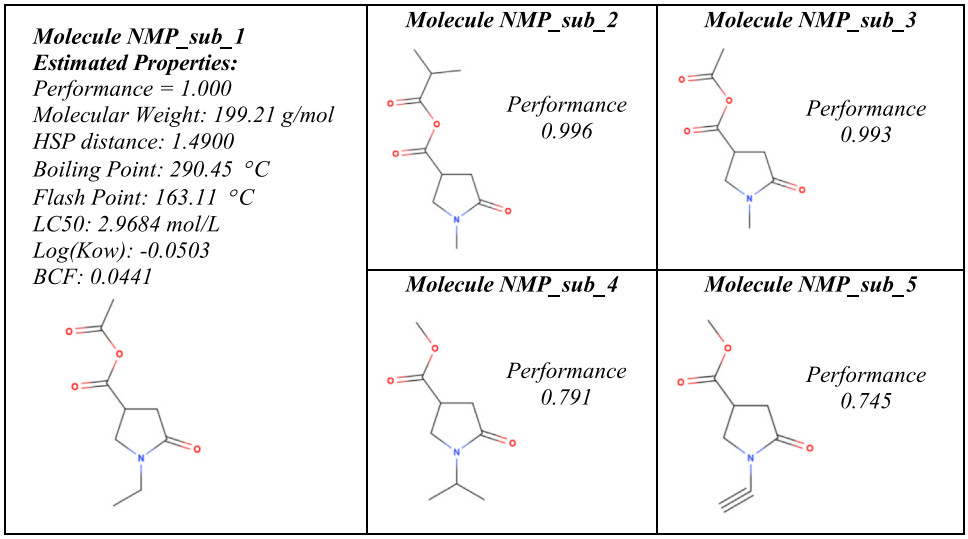

Fig. 7. Five candidates to replace aprotic highly dipolar solvent NMP with biosourced fragment NACP.

parameters are the population size (100), elitism (10), crossover (P=65), mutation (P=15), insertion (P=10), deletion (P=10). Also, the formation of cyclic and aromatic compounds is not allowed.

The result output file displays a list of a hundred molecules rated by their performance (Heintz et al., 2014). Three different CAMD searches are run, targeting the replacement of NMP, DMF and DMSO respectively, through a change of the HSP distance property target (see Table 1).

Figs. 7–9 show new molecules that have been identified by the IBSS software tool, for NMP, DMF and DMSO substitution case, respectively. The target properties, which are estimated by IBSS, are listed for the top molecule.

The strength of a CAMD approach is to explore the space of chemical structures that match the set of targeted properties very efficiently. However, results need clearly an expert eye with chemical background to validate their feasibility in terms of chemical synthesis at an industrial scale. We have partly covered that issue by imposing the NACP bio-based fragment which can be readily synthesized in an eco-friendly manner. However with substituents R1 and R2, the new structures may not be realistic. For example, candidates NMP_sub_1,

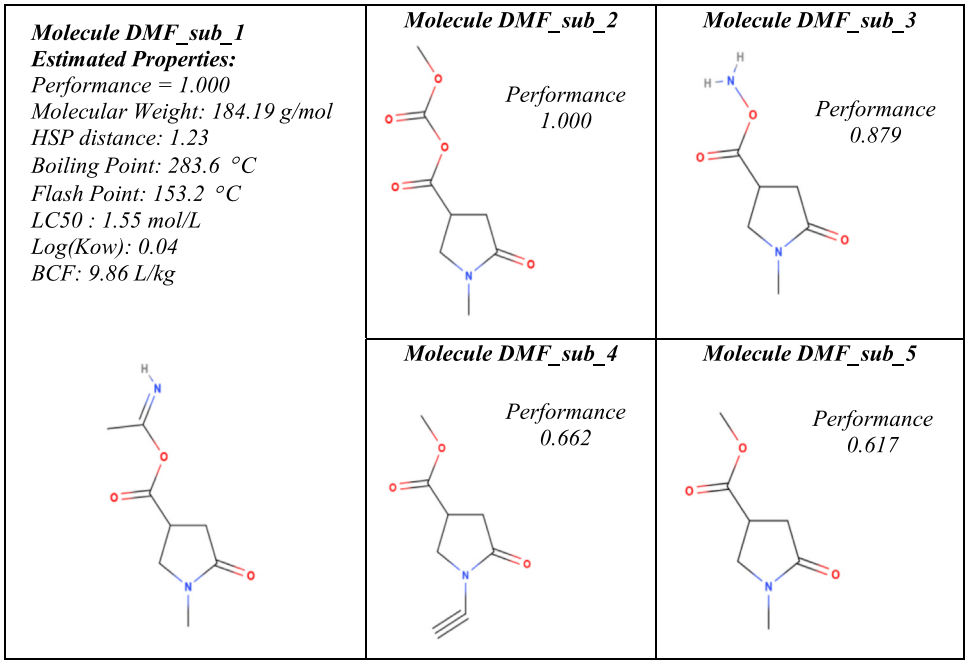

Fig. 8. Five candidates to replace aprotic highly dipolar solvent DMF with biosourced fragment NACP.

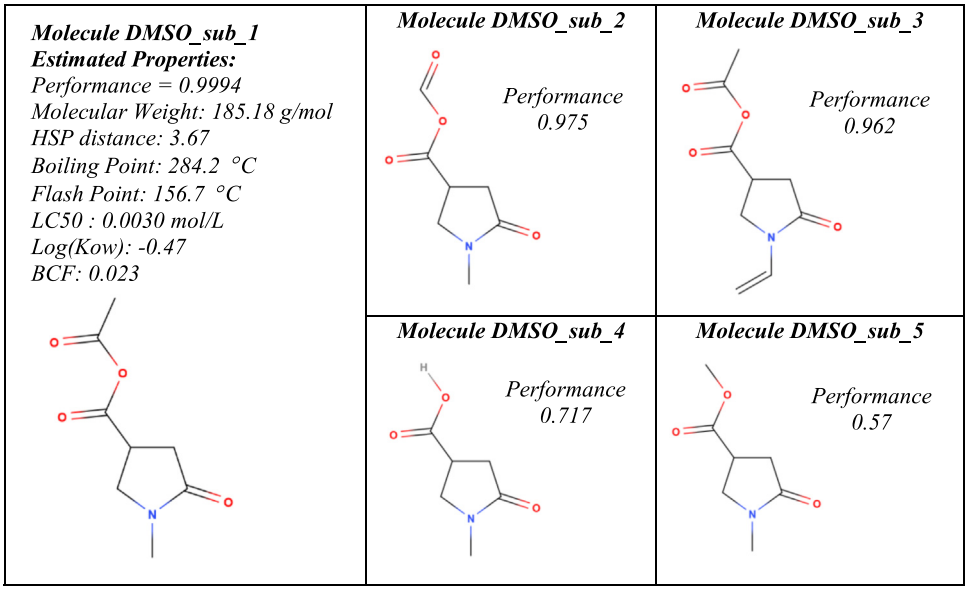

Fig. 9. Five candidates to replace aprotic highly dipolar solvent DMSO with biosourced fragment NACP.

2 and 3 in Fig. 7 are anhydride acids which would be readily hydrolyzed in aqueous medium or would react with alcohols or amines to give esters or amides respectively. Hence their use is not recommended in cosmetics where such reactions could occur. Molecules NMP_sub_4 and 5 with aliphatic substituents R1 and R2 are more stable. Their performance is lower but is still above that of NACP alone that reaches 0.579 (Table 3). This demonstrates the ability of IBSS to enhance the properties of GRASS lead compounds and the relevance of the framework.

For DMF substitutes (Fig. 8), a similar analysis can be made. Among the most performant candidates, structures with nitrogen and oxygen groups occurs, synthesis of which might be possible but is not straightforward. Using aliphatic radicals lowers significantly the performance that gets close to that of the NACP alone that reaches 0.619 (Table 3) for substituting DMF. DMF_Sub_4 structure is the same as NMP_sub_5, with a different performance because the HSP target values of DMF and NMP are different.

For DMSO substitutes (Fig. 9), the same anhydride acid structure as that of NMP_sub_3, shows the best performance, but the problem of poor stability exist for such structures. For DMSO_sub_1, because its HSP distance is 3.67 which is greater than the target (3.6), its performance is not equal to one. The simplest NACP derivatives with alkyl substituents such as DMSO_sub_5, exhibit a performance below that of NACP alone (0.885 in Table 3).

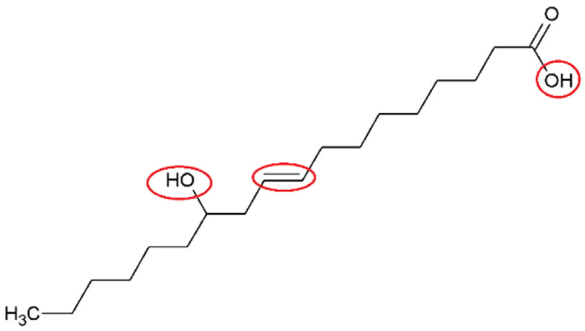

Fig. 10. Ricinoleic acid with the sites to functionalization.

3.2.4. Aprotic highly dipolar solvent choice phase

The choice phase requires to update the candidate molecule performance with real property values. Most of the compounds above are not known. Hence experiments would be required, in particular for validating their solubilization power as a solvent substitute

The second stage of the choice phase consists in determining if, in light of the experimental results from laboratory, the molecule alternative tested is still considered satisfactory, or if another alternative from the set shall be tested. Similarly to the first stage, another possible outcome is to go back to the intelligence and the design phases for obtaining a new set of alternatives. For the present case study, the prediction of EHS properties is based on usual LC50, BCF and log K_{ow} values. But they are poorly discriminant for all NACP derivatives. Reasons could be that either all NACP derivative have low EHS impacts as claimed in the literature (Richard and Muller, 2010) or that the methods are not state of the art. Indeed new methods which are more accurate and developed over a larger set of data have been recently published (Hukkerikar et al., 2012b) and could be used instead.

A posteriori, we can assess the sustainability degree of the solution over six dimensions as suggested by Gagnon et al. (2012), with sustainability shade ranging from A (minimal) to D (state of the art). The 'design process' dimension consists in counting how many tasks were covered among the 22 tasks listed by Gagnon in his integrated sustainable design process. With 13 tasks, including all tasks listed as critical by Gagnon (see Section 2.5), the 'design-process' dimension is graded 'B-shade'. The 'sustainability issue covered' dimension is graded "B-shade" as issues covering partially all three sustainability pillars are considered. The 'indicator relevance' dimension should get an excellent 'D-shade' grade since a systematic search of bio-sourced solvent was run. But it is degraded to a poor grade 'A-shade' since the initial EHS index models were found unsuitable. The 'analysis tool accuracy' dimension is graded a 'B-shade' because most property estimation models but the EHS index methods are near state of the art predictive models. The 'alternative performance' dimension is graded 'D-shade' because novel bio-based solvents have been found, reducing the impacts of existing solvents such as NMP or DMF. The 'decision-making' dimension is graded 'C-shade' because all sustainability pillars are addressed and so in a balanced manner. Overall with one A, two Bs, one C and one D-shade, the experts estimate that the product development process is reasonably sustainable but can be improved. Thus the decision is taken to run again the CAPD tool by selecting new property estimation methods for the EHS impacts assessment as the 'indicator relevance' dimension is responsible for the current worst 'A-shade'. This new round is not presented.

3.3. Case study 2: biolubricants from vegetable oils

Presentation of this case study is reduced and the decision-making process sketched in Fig. 1 is not run in full details. Besides, running GRASS to generate so-called lead compound is not necessary because a bio-sourced lead molecule exist, namely ricinoleic acid.

3.3.1. Biolubricant intelligence phase

The commodity plant oils have been studied and used as sources of improved chemicals, such as paint formulations, coatings, lubricants, cutting-fluids, printing inks, solvents and plastics (Teles dos Santos et al., 2014). Such feedstocks are composed mainly by triglycerides formed by three fatty acids, with different chain lengths and degree of unsaturations. In castor oil, the main fatty acid is ricinoleic acid ((9Z, 12R)-12-hydroxy-9-octadecenoic acid), which can be functionalized into different molecular structures. Experimental studies have synthetized biolubricants from ricinoleic acid (Salimon et al., 2012; Yao et al., 2010) and measured their properties afterwards. They have concluded that there is a need to deal with the trade-off between opposing properties and to explore a greater variety of molecular structures. Thus, the use of CAMD can be useful to find that trade-off. Fig. 10 shows the ricinoleic acid molecular structure with the sites where the chemical functionalization is normally done (carboxyl group, double bond and hydroxyl group).

3.3.1.1. Defining properties. Plant oils have a tendency to crystallize and this limits their use at low temperature operating conditions. An improvement in the cold flow behavior is desired and melting point is used to evaluate candidate molecules. Transportation and storage safety is another concern when designing new chemicals and flash point is evaluated as well. The lubricity plays a key role and the viscosity is used in the objective function. Finally, environmental concerns are evaluated using the potential to bioaccumulation by means of octanol–water partition coefficient. The resistance to oxidative degradation and easiness of synthesis are concerns that are not adressed in the design phase here. Table 4 summarizes the physical properties to be satisfied by a potential candidate to be used as a biolubricant.

3.3.1.2. Building blocks. This work uses two strategies to generate ricinoleic acid derivatives: the first one adds groups to the double bond site and to the carboxyl group (Fig. 11a). The carboxyl group is normally esterified in experimental studies with different alcohols, leading to esters with different chain lengths and properties. The second strategy uses the hydroxyl group and the carboxyl group to generate derivatives, the double bond remaining unchanged (Fig. 11b). The combinatorial nature of the problem leads to an extremely large number of possible molecules, difficult to handle experimentally.

Table 4
Real and calculable properties with corresponding CAMD parameters.

Real property	Calculable property	Target value	Prediction method	Performance function (Eq. (2))	Weight
Liquid at usage temp and good cold flow behavior	Melting point	< -10 °C	HSKASG2012	G (10;0.8)	1
Non flammability at high temp	Flash point	> 290 °C	HSKASG2012	G (5;0.8)	1
Viscous	Viscosity	> 0.28 Pa s@25 °C	JR1987	G (10;0.8)	1
Low potential to bioaccumulation	log K_{ow}	< 3	HSKASG2012	G (2;0.8)	1

HSKASG2012:Hukkerikar et al. (2012a). JR1987:Joback and Reid (1987).

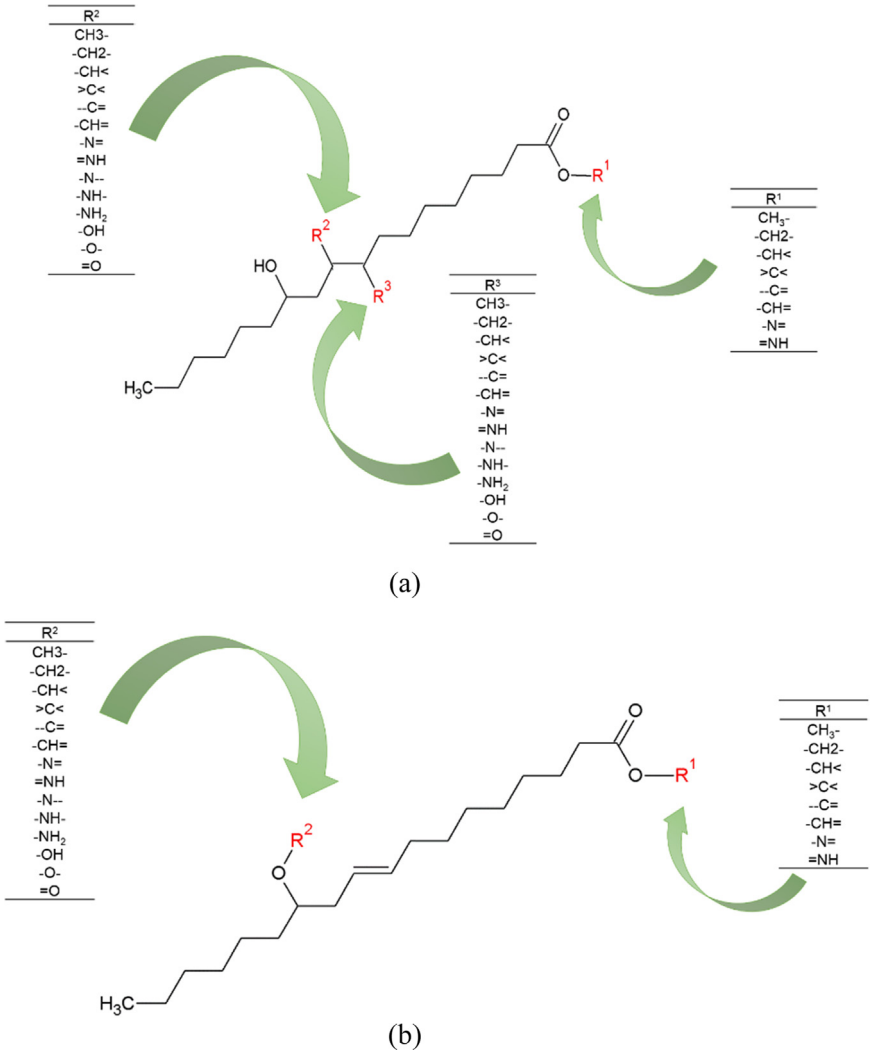

Fig. 11. Building blocks used to generate ricinoleic acid derivatives with two different strategies.

3.3.2. Biolubricant design phase

3.3.2.1. CAMD parameters. In addition to the molecule structure, building blocks and targeted properties, search algorithm parameters are sets. Several tests were performed using different parameters: number of generations (up to 300), size of the population (100, 50 and 30), elitism (10, 5 and 3) and maximum number of building blocks kmax allowed to be added to the ricinoleic acid (from 1 to 10). The reported results refers to the following set of parameters: number of generations (300), population size (50), elitism (5), crossover (P=65), mutation (P=15), insertion (P=10), deletion (P=10). Also, the formation of cyclic and aromatic compounds is not allowed, although they are possible with the IBSS search algorithm.

3.3.2.2. Results. In general, the search algorithm had to deal with two opposing properties: the larger the molecule, the larger the viscosity but with larger melting point and larger log K_{ow}. The hydroxyl group seemed to improve the performance of the generated molecules and solutions with kmax larger than 5 were seldom found. Table 5 shows some solutions found by the method for the two strategies adopted.

All molecules have predicted melting points greater than the target. Indeed, this is one of the main challenges of the problem: increase viscosity without increasing melting point. These molecules must be further analyzed regarding their melting point. For the majority of the generated molecules, the flash point and viscosity are within the desired range while melting points is larger than the target and the octanol–water partition coefficient is slightly greater than the target.

Table 5
Molecular fragments, performance and predicted properties for commonly found solutions (300th generation, kmax=5).

Strategy A	R1	R2	R3	Performance	*Tm* (°C)	F_p (° C)	Viscosity (Pa s)@25 °C	Log K_{ow}
Molecule 1	$-CH_3$	$-OCH_2OH$	-CH(OH)OH	0.75	145.49	448.65	111.85	2.85
Molecule 2	$-CH_2CH_3$	–CH(OH)OH	–OH	0.74	143.85	430.48	104.96	3.56
Molecule 3	$-CH_3$	–OH	–OH	0.74	116.59	345.85	7.98	3.88
Molecule 5	$-CH_3$	$-OCH_2OH$	–OH	0.74	121.35	375.48	9.98	3.63
Molecule 4	$-CH_3$	–CHO	–OH	0.73	118.49	325.29	1.91	4.16
Strategy B								
Molecule 1	$-CH_2CH_3$	–CH(OH)OH	–	0.68	101.87	295.75	1.02	5.48
Molecule 2	$-CH(CH_3)CH_3$	–CH(OH)OH	–	0.66	102.05	299.71	1.18	5.83
Molecule 3	$-CH_2CH_2CH_3$	–CH(OH)OH	–	0.65	105	307.16	1.19	5.93
Molecule 4	$-CH_2CH_2CH_3$	$-CH_2CH_2OH$	–	0.40	49.83	276.31	0.14	7.03
Molecule 5	$-CH_3$	$-CH_3$	–	0.37	25.43	175.46	0.009	6.93

The IBSS tool saves the molecular structures generated and the corresponding calculated properties for any given generation in separate files, which can be further used to screen the thousands of generated molecules.

3.3.2.3. Biolubricant choice phase. The selection of a viable alternative during the choice phase is not conclusive because of the comments about the melting point and viscosity opposing trends.

As future work, the authors suggest three improvements: i) the automatic search procedure could integrate more accurate GC methods specific for lipids compounds, aiming to be more accurate in melting point predictions; ii) more complex building blocks could be used to generate ricinoleic derivatives and iii) the use of mixtures of fatty acids derivatives could be evaluated in order to improve viscosity while keeping low melting point. Concerning melting point, several experimental works have demonstrated the formation of eutectic points in fatty systems, leading to mixtures with a melting point lower than that of the pure compounds (Costa et al., 2007; Robustillo et al., 2013). This indicates that mixtures of fatty acids derivatives can be also considered when searching for biolubricants candidates. In this case, composition and chemical structure of mixture's compounds are decision variables. Hence the experts would suggest a new round of search starting at the design phase for finding mixtures. This will be run in the future.

4. Conclusions

Considering the growing importance for new products to meet ever stricter Environmental, Health and Safety specifications while matching functional properties, an original way is to design new chemical products that can be traced to renewable material stocks. Another way is to bias the search towards new molecules that can be synthesized according to green chemistry principles.

We have presented in this paper an integrated decision-making process that combines the use of computer-aided organic synthesis and computer aided molecular design approaches to address that issue in both ways. By involving all stakeholders across the enterprise hierarchy during the so-called intelligence phase, we are able to build a requirement tree for the molecules with the help of a facilitator. Thanks to the use of standard modeling languages SBVR for business rules and OCL for constraints, requirements are stated unambiguously and match all stakeholders perception of the ideal molecule to be found. The requirement tree is then processed in the design phase to define input files for running computer aided organic synthesis CAOS and molecular design CAMD tools. The CAOS provides lead compounds starting from a renewable feedstock, exemplified with the itaconic acid feedstock in Moity et al. (2014). These first generation molecules display an overall performance that is usually not satisfactory. Then the CAMD step aims at enhancing the properties by proposing derivatives thanks to additional substitutes. Hence the CAOS + CAMD combination proposes candidate molecules matching desired properties, with a possible origin from renewable feedstocks and with a possible eco-friendly synthesis pathway that was imposed in the CAOS step. Ultimately we evaluate the sustainability of the whole decision-making process and results. It is noticed that the prediction methods for EHS impacts are not satisfactory. The developed CAMD tool (IBSS) was also used to generate ricinoleic acid derivatives aiming to find new biolubricants that could be produced from a biomass feedstock (castor oil).

We have shown the pertinence of running a decision-making process with all stakeholders across the enterprise for finding new molecules. These molecules can be claimed as sustainable according to their low EHS impacts, but also because thanks to the proposed methodology, they are partly originated in structure and in their synthesis process to renewable materials. However, some challenges remain, especially in the development of more accurate property prediction models, in particular for bio-sourced molecules.

References

Achenie, L.E.K., Gani, R., Venkatasubramanian, V., 2003. Computer Aided Molecular Design: Theory and Practice. Elsevier, Amstersdam.
Anastas, P.T., Warner, J., 1998. Green Chemistry Theory and Practice. Oxford University Press, Oxford, p. 135.
Anastas, P.T., Zimmerman, J.B., 2003. Design through the 12 principles of green engineering. Environ. Sci. Technol., 94–101.
Azapagic, A., Millington, A., Collett, A., 2006. A methodology for integrating sustainability considerations into process design. Chem. Eng. Res. Des. 84, 439–452.
Azapagic, A., Perdan, S., 2005. An integrated sustainability decision-support framework - Part I: problem structuring. Int. J. Sustain. Dev. World Ecol. 12, 98–111.
Bandres, M., de Caro, P., Thiebaud-Roux, S., Borredon, M.-E., 2011. Green syntheses of biobased solvents. Comptes Rendus Chimie 14, 636–646.
Barone, R., Chanon, M., Vernin, G., Parkanyi, C., 2005. Generation of potentially new flavouring structures from thiamine by a new combinatorial chemistry program. In: Spanier, A.M., Shahidi, F., Parliment, T.H., Mussinan, C., Tratras Contis, E. (Eds.), Food Flavor and Chemistry; Explorations into the 21st Century 2005. RSC, Cambridge, UK, pp. 175–212.
Barone, R., Chanon, M., Vernin, G., Parkanyi, C., 2010. Computer-aided organic synthesis as a tool for generation of potentially new flavoring compounds from ascorbic acid in recent advances in food and flavor chemistry: food flavors and encapsulation. In: Ho, C.T., Contis, E.T., Mussinan, C.J., Shahidi, F. (Eds.), Health Benefits, Analytical Methods, and Molecular Biology of Functional Foods 2010. RSC, Cambridge, UK, pp. 81–127.
BASF, Mixtures of itaconic acid or itaconic acid derivatives and primary amines for producing 1,3- and 1,4-alkyl methyl pyrrolidones. Patent WO2010063617, 2009.
Bergez-Lacoste, M., Thiebaud-Roux, S., De Caro, P., Fabre, J.F., Gerbaud, V., Mouloungui, Z., 2014. From chemical platform molecules to new biosolvents: design engineering as

substitution methodology. Biofuels Bioprod. Biorefin. 8, 438–441.
Bruntland report, 1987. “our common future”. ⟨http://www.un-documents.net/our-common-future.pdf⟩. (Last accessed march 2016).
Catoire, L., Paulmier, S., Naudet, V., 2006. Experimental determination and estimation of closed cup flash points of mixtures of flammable solvents. Process Saf. Prog. 25 (1), 33–39.
Conte, E., Gani, R., Ng, K.M., 2011. Design of formulated products: a systematic methodology. AIChE J. 57, 2431–2449.
Conte, E., Gani, R., 2011. Chemicals-based formulation design. Comput. Aided Chem. Eng. 29, 1588–1592.
Costa, R., Moggridge, G.D., Saraiva, P.M., 2006. Chemical product engineering: an emerging paradigm within chemical engineering. AIChE J. 52, 1976–1986.
Costa, M.C., Rolemberg, M.P., Boros, L.A.D., Krähenbühl, M.A., de Oliveira, M.G., Meirelles, A.J.A., 2007. Solid–liquid equilibrium of binary fatty acid mixtures. J. Chem. Eng. Data 52, 30–36.
Dalkey, N., Helmer, O., 1963. An experimental application of the Delphi method to the use of experts. Manag. Sci. 9, 458–467.
Durand, M., Molinier, V., Kunz, W., Aubry, J.-M., 2011. Classification of organic solvents revisited by using the COSMO-RS approach. Chem. A Eur. J. 17, 5155–5164.
ECHA. REACH market final report 2012, (see ⟨http://ec.europa.eu/enterprise/sectors/chemicals/files/reach/review2012/market-final-report_en.pdf⟩). (last accessed october 2015).
Frenkel, M., Chirico, R.D., Diky, V.V., Dong, Q., Frenkel, S., Franchois, P.R., Embry, D.L., Teague, T.L., Marsh, K.N., Wilhoit, R.C., 2004. ThermoML – an XML-based approach for storage and exchange of experimental and critically evaluated thermophysical and thermochemical property data. 3. Critically evaluated data. J. Chem. Eng. Data 49, 381–393.
Gagnon, B., Leduc, R., Savard, L., 2012. From a conventional to a sustainable engineering design process: different shades of sustainability. J. Eng. Des. 23, 49–74.
Gani, R., 2004. Chemical product design: challenges and opportunities. Comput. Chem. Eng. 28, 2441–2457.
Gani, R., Brignole, E.A., 1983. Molecular design of solvents for liquid extraction based on UNIFAC. Fluid Ph. Equilib. 13, 331–340.
Hansen, C.M., 2007. Hansen Solubility Parameters. CRC press, Taylor & Francis Group, Boca Raton.
Harper, P.M., Gani, R., 2000. A multi-step and multi-level approach for computer aided molecular design. Comput. Chem. Eng. 24, 677–683.
Heintz, J., Belaud, J.P., Gerbaud, V., 2014b. Chemical enterprise model and decision-making framework for sustainable chemical product design. Comput. Ind. 65, 505–520.
Heintz, J., 2012. Systemic approach and decision process for sustainability in chemical engineering: application to computer aided product design. Thèse de doctorat. Institut National Polytechnique de Toulouse. ⟨http://ethesis.inp-toulouse.fr/archive/00002022/⟩.
Heintz, J., Belaud, J.P., Pandya, N., Teles dos Santos, M., Gerbaud, V., 2014a. Computer aided product design tool for sustainable chemical product development. Comput. Chem. Eng. 71, 362–376.
Herring III, R.H., Eden, M.R., 2015. Evolutionary algorithm for de novo molecular design with multi-dimensional constraints. Comput. Chem. Eng. 83, 267–277.
Hill, M., 2009. Chemical product engineering—the third paradigm. Comput. Chem. Eng. 33, 947–953.
HSPIP V3.0. 2010 Y-MB model from ⟨http://www.pirika.com/NewHP/Y-MB/Y-MB.html⟩.
Hukkerikar, A.S., Kalakul, S., Sarup, B., Young, D.M., Sin, G., Gani, R., 2012b. Estimation of environment-related properties of chemicals for design of sustainable processes: development of group-contribution+ (GC+) property models and uncertainty analysis. J. Chem. Inf. Model. 52 (11), 2823–2839.
Hukkerikar, A.S., Sarup, B., Ten Kate, A., Abildskov, J., Sin, G., Gani, R., 2012a. Group-contribution+ (GC+) based estimation of properties of pure components: Improved property estimation and uncertainty analysis. Fluid Ph. Equilib. 321, 25–43.
Hung, H.F., Kao, H.P., Juang, Y.S., 2008. An integrated information system for product design planning. Exp. Syst. Appl. 35 (1–2), 338–349.
Joback, R., Reid, R.C., 1987. Estimation of pure-component properties from group-contributions. Chem. Eng. Commun. 57, 233–243.
Karunanithi, A.T., Achenie, L.E.K., Gani, R., 2005. A new decomposition-based computer-aided molecular/mixture design methodology for the design of optimal solvents and solvent mixtures. Ind. Eng. Chem. Res. 44, 4785–4797.
Korichi, M., Gerbaud, V., Floquet, P., Meniai, A.-H., Nacef, S., Joulia, X., 2008. Computer aided aroma design I-Molecular knowledge framework. Chem. Eng. Process.: Process Intensif. 47, 1902–1911.
Marrero, J., Gani, R., 2001. Group-contribution based estimation of pure component properties. Fluid Ph. Equilib. 183–184, 183–208.
Marrero, J., Gani, R., 2002. Group-contribution-based estimation of octanol/water partition coefficient and aqueous solubility. Ind. Eng. Chem. Res. 41, 6623–6633.
Martin, T.D., Young, D.M., 2001. Prediction of the acute toxicity (96-h LC50) of organic compounds to the fathead minnow (pimephales promelas) using a group contribution method. Chem. Res. Toxicol. 14, 1378–1385.
Mattei, M., Hill, M., Kontogeorgis, G.M., Gani, R., 2014a. A comprehensive framework for surfactant selection and design for emulsion based chemical product design. Fluid Ph. Equilib. 362, 288–299.
Mattei, M., Yunus, N.A., Kalakul, S., Kontogeorgis, G.M., Woodley, J.M., Gernaey, K.V., Gani, R., 2014b. The virtual product–process design laboratory for structured chemical product design and analysis. Comput. Aided Chem. Eng. 33, 61–66.
Moity, L., Durand, M., Benazzouz, A., Pierlot, C., Molinier, V., Aubry, J.-M., 2012. Panorama of sustainable solvents using the COSMO-RS approach. Gr. Chem. 14, 1132–1145.
Moity, L., Molinier, V., Benazzouz, A., Barone, P., Marion, P., Aubry, J.-M., 2014. In silico generation of bio-based commodity chemicals: application to the design of itaconic acid-based solvents. Gr. Chem. 16, 146–160.
Moity, L., Molinier, V., Benazzouz, A., Joossen, B., Gerbaud, V., Aubry, J.-M., 2016. A “top-down” in silico approach for designing ad hoc bio-based solvents: application to glycerol-derived solvents of nitrocellulose. Gr. Chem. . http://dx.doi.org/10.1039/c6gc00112b
Murray-Rust, P., Rzepa, H.S., 2001. Chemical markup, XML and the World-Wide Web. 2. information objects and the CMLDOM. J. Chem. Inf. Comput. Sci. 41, 1113–1123.
Ng, K.M., 2004. MOPSD: a framework linking business decision making to product and process design. Comput. Chem. Eng. 29, 51–56.
Nieddu, M., Garnier, E., Bliard, C., 2012. The emergence of doubly green chemistry, a narrative approach. Eur. Rev. Ind. Econ. Policy, 1.
OMG, 2006. Object Constraint Language OMG Available Specification Version 2.0 OMG, Needham, MA, [WWW Document] available at: ⟨http://www.omg.org/spec/OCL/2.3.1/PDF/⟩.
OMG, 2008. Semantics of Business Vocabulary and Business Rules (SBVR), v1.0. OMG, Needham, MA, [WWW Document] available at: ⟨http://www.omg.org/spec/SBVR/1.0/PDF/⟩.
Papadopoulos, A.I., Stijepovic, M., Linke, P., Seferlis, P., Voutetakis, S., 2013. Molecular design of working fluid mixtures for organic rankine cycles. Comput. Aided Chem. Eng. 32, 289–294.
REACH, 2006. ⟨http://echa.europa.eu/en/regulations/reach⟩. (Last accessed march 2016).
Richard, H., Muller, B., 2010. L'Oreal, Ester Solvents Derived from 4-carboxy-2-pyrrolidinone Formulated into UV-screening Cosmetic Compositions. France Patent US 2010/0143276 A1.
Robustillo, M.D., Barbosa, D.F., Meirelles, A.J.A., Pessôa Filho, P.A., 2013. Solid–liquid equilibrium in ternary mixtures of ethyl oleate, ethyl laurate and ethyl palmitate. Fluid Ph. Equilib. 339, 58–66.
Salimon, J., Salih, N., Yousif, E., 2012. Biolubricant basestocks from chemically modified ricinoleic acid. Journal of King Saud University – Science 24, 11–17.
Samudra, A., Sahinidis, N.V., 2013. Optimization-based framework for computer-aided molecular design. AIChE J. 59, 3686–3701.
Simon, H.A., 1960. The New Science of Management Decision, The Ford Distinguished Lectures. Harper and Row, New York and Evanston.
Smith, B.V., Ierapepritou, M.G., 2010. Integrative chemical product design strategies: reflecting industry trends and challenges. Comput. Chem. Eng. 34, 857–865.
Solvason, C.C., Chemmangattuvalappil, N.G., Eden, M.R., 2009. A systematic method for integrating product attributes and molecular synthesis. Comput. Chem. Eng. 33 (5), 977–991.
Teles dos Santos, M., Gerbaud, V., Le Roux, G.A.C., 2014. Beyond biofuels: economic opportunities, recent advances and challenges in property modeling for vegetable oils. Gr. Process Synth. 3, 401–410.
UNGHS: United Nation Global Harmonization System of Classification recommendation (rev 5), 2013. ⟨http://www.unece.org/fileadmin/DAM/trans/danger/publi/ghs/ghs_rev05/English/ST-SG-AC10-30-Rev5e.pdf⟩.
Vaidyanathan, R., El-Halwagi, M., 1996. Computer-aided synthesis of polymers and blends with target properties. Ind. Eng. Chem. Res. 35, 627–634.
Veith, G.D., Konasewich, D.E., 1975. Structure–activity correlations in studies of toxicity and bioconcentration with aquatic organisms. In: Proceedings of the Symposium held in Burlington, Ontario, Canada 1975 and Computers and Chemical Engineering, 1999, 23, pp. 1381–1394.
Weis, D.C., Visco, D.P., 2010. Computer-aided molecular design using the signature molecular descriptor: application to solvent selection. Computers and Chemical Engineering. 34, 1018–1029.
Werpy, T., G. Petersen, G., 2004. Top Value Added Chemicals from Biomass: Volume I – Results of Screening for Potential Candidates from Sugars and Synthesis Gas, Report DOE/GO-102004-1992.
Yao, L., Hammond, E.G., Wang, T., Bhuyan, S., Sundararajan, S., 2010. Synthesis and physical properties of potential biolubricants based on Ricinoleic acid. J. Am. Oil Chem. Soc. 87 (8), 937–945.